PUBLICATIONS OF THE INDIAN METEOROLOGICAL DEPARTMENT.

The following is a list of the more important publications of the India Meteorological Department :—

The Indian Meteorologist's *Vade Mecum*, Part I, 2nd Edition. (1883) . . . Price Rs. 3	Henry F. Blanford.
The Indian Meteorologist's *Vade Mecum*, Part II (1877) . . . Price Rs. 5*	Ditto.
Instructions to Observers of the Indian Meteorological Department, 2nd Edition, (1902) . . . Price Rs. 3	Sir John Eliot.
Tables for the reduction of Meteorological Observations in India, 2nd Edition. (1889) . . . Price Rs. 2	Henry F. Blanford.
Handbook of Cyclonic storms in the Bay of Bengal for the use of sailors, 2nd Edition, Vol. I—Text. (1900). . . . Price Rs. 3	Sir John Eliot.
Handbook of Cyclonic storms in the Bay of Bengal for the use of sailors, 2nd Edition, Vol. II—Plates. (1901). Price Rs. 1-8	Ditto.
Cyclone Memoirs, Part I—Bay of Bengal Cyclone of May 20th to 28th, 1887. (1888) . . . Price Re. 1*	Ditto.
Cyclone Memoirs, Part II—Bay of Bengal Cyclone of August 21st to 28th, 1888. (1889) . . . Price Rs. 3	Ditto.
Cyclone Memoirs, Part II[illegible] 13th to 20th, and Oct[illegible] Cyclone of November[illegible]	Ditto.
Cyclone Memoirs, Part I[illegible] of storms in the Arabian Sea [illegible] history of all recorded storms in the Arabian Sea from 1848—1889. (1891) . . . Price Rs. 3	W. L. Dallas.
Cyclone Memoirs, Part V—Account of three Cyclones in the Bay of Bengal and Arabian Sea during November, 1891. (1893) . . . Price Rs. 3	Sir John Eliot.
Report of the Midnapore and Burdwan Cyclone of the 15th and 16th of October, 1874. (1875) . . . Price Rs. 3*	W. G. Willson.
Report of the Vizagapatam and Backergunge Cyclones of October, 1876. (1877) . . . Price Rs. 3*	Sir John Eliot.
Report on the Madras Cyclone of May, 1877. (1879) Price Rs. 3*	Ditto.
Monthly weather charts of the Bay of Bengal and adjacent sea north of the equator, showing mean pressure, winds and currents. (1886) . . . Price Rs. 5	Henry F. Blanford.
Monthly weather charts of the Arabian Sea and the adjacent portion of the North Indian Ocean, showing mean pressure, winds and currents. (1888) . . . Price Rs. 5	Sir John Eliot.
Charts of the Bay of Bengal and adjacent sea north of the equator, showing the specific gravity, temperature and currents of the sea surface (1887) . . . Price Rs. 1-8	W. L. Dallas.
Climatological Atlas of India (1906). Price Rs. 27 or 36 shillings	Sir John Eliot.
Meteorological Atlas of the Indian seas and the North Indian Ocean (1908) . . . Price Rs. 13 or 13 shillings 6 pence	W. L. Dallas.
Daily weather reports and charts of the Indian monsoon area for the years 1893 to 18[illegible] . . . Price, each month, Re. 1*	Departmental.
Normal weather or pilot charts of the Indian monsoon area for 8 A.M. for each month, November 1900 to August 1903† . . . Price, each month, Annas 4	Ditto.
Rainfall data of India for the years 1891—1906 (16 volumes). . . . Price for each year, Rs. 5‡	Ditto.
R[illegible] India for the years 1875—1891 . . . Price each, Re. 10	Ditto.
In[illegible] Summaries for the years 1891—1906 (16 parts) § . . . Price [illegible]	
Monthly Weather Reviews for each m[illegible] July, 1903§	

Indian Meteorological Memoirs, Vol. I, cont.

Part I, comprising—	
On the winds of Calcutta—An analysis of ten years [illegible] observations of the wind vane and four years' anemograms	
The meteorology and climate of Yarkhand and Kashgar, being chiefly a discussion of registers kept by Dr. J. Scully in 1874-75	Henry F. Blanford.
The diurnal variation of the barometer at Simla. . . . Price Rs. 3*	
Part II, comprising—	
Storms in Bengal during the year 1876 accompanied with increased atmospheric pressure and the apparent reversal of the normal diurnal oscillation of the barometer . . .	Sir John Eliot.
On the rainfall of Benares considered in relation to the prevailing winds . . .	S. A. Hill.
On the diurnal variation of the barometer at Indian stations, Part I, Calcutta and Hazaribagh . . . Price Rs. 3*	Henry F. Blanford.
Part III, comprising—	
Variation of rainfall in Northern India . . .	S. A. Hill.
Meteorological and hypsometrical observations in Western Tibet, recorded by Dr. J. Scully, with a discussion . . . Price Rs. 3	Henry F. Blanford.
Part IV, comprising—	
The winds of Kurrachee . . . Price Rs. 6	Fred. Chambers.
Part V, comprising—	
Some results of the meteorological observations taken at Allahabad during the ten years 1870—79 . . .	S. A. Hill.
The diurnal variation of the barometer at Indian stations, Part II, Goalpara, Patna and Leh . . . Price Rs. 3	Henry F. Blanford.
Part VI, comprising—	
The Meteorology of the North-West Himalaya . . . Price Re. 1	S. A. Hill.

Indian Meteorological Memoirs, Vol. II, containing :—

Part I, comprising—	
Account of south-west monsoon storm of the 12th to the 24th of September, 1878, in the north of the Bay of Bengal . . .	Sir John Eliot.
List of cyclones on the West Coast of India and in the Arabian Sea up to the end of year 1881 . . . Price Rs. 2	Fred. Chambers.
Part II, comprising—	
Note on the foregoing list of cyclones and on the Gujarat land cyclone of July 11th to 13th, 1881 . . .	Henry F. Blanford.
On the temperature of North-Western India . . . Price Rs. 2	S. A. Hill.
Part III, comprising—	
Account of south-west monsoon storms of the 5th to the 19th October, 1882, in the Bay of Bengal . . . Price Rs. 2	Sir John Eliot.
Part IV, comprising—	
Account of the south-west monsoon storms generated in the Bay of Bengal during the years 1877 to 1881 . . . Price Rs. 2	Ditto.
Part V, comprising—	
Observations of temperature and humidity at a height of 40 feet above the ground at Alipore Observatory, Calcutta . . . Price Re. 1	S. A. Hill.
Indian Meteorological Memoirs, Vol. III, containing a full discussion of the rainfall of India and cognate subjects, complete in 4 parts . . . Price each part Rs. 3*	Henry F. Blanford.

Indian Meteorological Memoirs, Vol. IV, containing :—

Part I, comprising—	
Account of the south-west monsoon storm of the 12th to the 17th of May, 1884, in the Bay of Bengal and at Akyab . . .	Sir John Eliot.
On the diurnal variation of the rainfall at Calcutta . . .	Henry F. Blanford.
The meteorological features of the southern part of the Bay of Bengal . . . Price Rs. 3*	W. L. Dallas.
Part II, comprising—	
The False Point cyclone of September 22nd, 1885 . . . Price Rs. 3*	Sir Alexander Pedler.
Part III, comprising—	
On the ground temperature observations made at the old Observatory, Allahabad . . . Price Rs. 1-8*	S. A. Hill.
Part IV, comprising—	
List and brief account of the south-west monsoon storms generated in the Bay of Bengal during the years 1882 to 1886 . . . Price Rs. 3	Sir John Eliot.
Part V, comprising—	
The cyclonic storms of November and December, 1886, in the Bay of Bengal . . .	Ditto.
The cyclone of the 25th May to the 2nd June, 1881, in the Arabian Sea . . . Price Rs. 3	Fred. Chambers.
Part VI, comprising—	
On temperature and humidity observations made at Allahabad at various heights above the ground . . . Price Rs. 1-8	S. A. Hill.
Part VII, comprising—	
The Arabian Sea cyclone of the 4th to the 13th June, 1887 . . .	Fred. Chambers.
On the Meteorology and Climatology of Northern Afghanistan . . . Price Rs. 1-8*	W. L. Dallas.
Part VIII, comprising—	
An account of the more important cold weather storms in India during the years 1876 to 1891 . . . Price Rs. 3*	Sir John Eliot.

* Copies of publications to the price of which an asterisk is appended are out of print.

† The normal weather or pilot chart of the Indian monsoon area for 8 A.M. is a monthly publication. Copies for June, August and September, 1901, and January [illegible] int.

‡ [illegible] 1876, 1878 to 1881, 1884, 1887 and 1890 are out of print.

[illegible] he years 1891 to 1904 are out of print.

§ [illegible] for the month. Copies of Monthly Weather Review for 1891—97 [illegible] to [illegible] January, 1904, are out of print.

Continued on page iii of this cover.

VI.—CORRELATION IN SEASONAL VARIATION OF CLIMATE.

BY

Gilbert T. Walker, M.A., Sc.D., F.R.S.

INTRODUCTION.

A cursory examination of the seasonal variations of any country will show that some of the departures from average conditions are directly related to other abnormal features, and the method by which the effects are produced is often fairly well known, as is the case when diminution of pressure is due to rise of temperature. Other departures, however, are connected with abnormal features in distant parts of the earth, and difficulties are experienced in ascertaining not only the nature of the results due to any variation, but also the chain of causes and effects by which the results are produced; as examples of the latter may be quoted the favourable influence upon Indian monsoon rainfall of the conditions which produce high temperature in the interior of Australia or high pressure in the Argentine Republic. Before attempting therefore to investigate the phenomena on physical lines it appears desirable to ascertain by purely empirical methods the character of as many relationships as possible in the hope of being able to pick out from the results so obtained a number of which the physical explanation is clear. If we can in this way find the intermediate links in the chain of causes we may replace an intricate problem by a number of simpler ones.

2. Empirical methods may be divided into graphical and numerical, and of these the first are open to some objection. For although curves showing the changes of magnitude of two closely related quantities will make their connection obvious, such graphical methods are far from satisfactory when the disturbing factors are numerous or the connection sought is slight. The same curves have been interpreted in opposite senses by different authors and in such cases numerical methods like those of statistics seem inevitable; while in nearly all cases they appear desirable, inasmuch as they give quantitative instead of qualitative results and are free from subjective influences. They also afford a criterion of the reliability of a computed relationship by comparing it with the probable amount of fictitious relationship which we may expect to be produced by mere accident even in entirely independent factors.

3. It must be remembered that the number of years for which reliable data are available over a large part of the earth does not exceed thirty, and that some of the climatic elements with which we are concerned, especially rainfall and solar activity as measured by sunspot numbers, probably undergo larger percentage variations, than do the quantities to which statistical methods are usually applied. The deviations from the exponential law of distribution may also be appreciable; and as the proof ordinarily given of the formula for the correlation coefficient is, strictly speaking, only applicable to cases in which the exponential law of distribution applies, it appears desirable to seek some justification which is as free

as possible from hypotheses as to the frequency of occurrence of the variations of the magnitudes concerned[1].

4. Let us consider two series of n quantities each, $X_1, X_2 \ldots\ldots X_n$ and $Y_1, Y_2 \ldots\ldots Y_n$; let them be associated in pairs, X_1 with Y_1, X_2 with Y_2 &c., and let their departures from their respective average[2] values be $x_1, x_2, \ldots\ldots x_n$ and $y_1\ y_2, \ldots\ldots y_n$, so that $Sx=0$, $Sy=0$, where Sx stands for the sum of the n terms $x_1, x_2 \ldots\ldots x_n$, and similarly $Sy=y_1+y_2+\ldots\ldots+y_n$. A primitive way of ascertaining the extent to which the values of the terms of the one series are affected by those of the other is that of counting the number of times p in which the values of x have the same sign as those of y, and the number of times q in which the signs are different. It is obvious that when the relationship is close the values of x will in most cases have the same sign as those of y if X, Y tend to vary in the same direction, or the opposite sign if X, Y tend to vary in opposite directions. If there is very slight relationship the variations will be nearly independent and p/q will be nearly equal to unity. Thus the fraction $(p-q)/(p+q)$ of which the numerical value must lie between unity and zero, might be used to give a rough idea of relationship.

5. When the number n in the series is small this method will be inaccurate because it does not take into account the magnitude of the departures from normal. If, for example, five pairs of values of x and y be—

x	+1·12	−0·03	−0·94	+0·02	−0·17
y	+0·55	+0·04	−0·45	−0·13	−0·01

it will be seen that these indicate a direct relationship. For the values 0·01, 0·02, 0·03, 0·04 are so small as to afford no reliable indications, such small quantities being completely masked by accidental causes: thus the values of x tend to be about twice those of y. There are also three signs alike and two unlike, so that $(p-q)/(p+q)$ is $+1/5$. If however we consider a different case in which the same numbers occur in a different arrangement—

x	+1·12	−0·03	−0·94	+0·02	−0·17
y	−0·45	−0·13	+0·55	+0·04	−0·01

it will be seen that the second, fourth and fifth pairs afford no reliable indications, and the first and third establish a relationship of the inverse character, X and Y tending to vary in opposite directions with values of x numerically about double those of y. The value of $(p-q)/(p+q)$ is still $+1/5$ and is misleading. Thus we cannot rely upon $(p-q)/(p+q)$ as giving even a rough indication; and the reason of its failure is that it takes only the signs of the departures into account, not their magnitudes.

[1] Certain authors (e.g. Yule 'On the theory of correlation' in the Journal of the Royal Statistical Society, Vol. LX, December 1897) obtain the correlation coefficient without explicit assumptions as to the frequency of distribution; but they adopt the method of least squares, and the justification of this depends on the law of distribution.

[2] The consideration of the effects introduced by using the 'mode' instead of the 'median' value is deferred for the present.

6. If the variations of X and Y be not independent, it is natural to regard the departures x of X as made up of a portion governed by y and a portion independent of y. If the values of y be small the portion of x determined by y may, if squares of small quantities be neglected, be taken as ky, where k is a constant independent of y: when y is not small the hypothesis that its effect is proportional to its magnitude appears the simplest which will approximately represent the facts. Thus we write :—

$$\left.\begin{array}{l} x_1=ky_1+d_1 \\ x_2=ky_2+d_2 \\ \cdots\cdots\cdots\cdots \\ x_n=ky_n+d_n \end{array}\right\} \quad \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \quad (1)$$

where d_1, d_2, . . . d_n are remainders representing those portions of x which are dependent on factors other than y, and these may, when y alone is considered, be treated as if they were accidental.

7. In order to determine k a conceivable plan would be to form two groups of equations, the first containing all those in which y is positive and the second those in which y is negative. We should, on adding the equations of each group, obtain two equations of the form

$$\left.\begin{array}{l} S_1x=kS_1y+S_1d \\ S_2x=kS_2y+S_2d \end{array}\right\}$$

where S_1, S_2 indicate summations over equations containing positive and negative values of y. Since $Sx=0$, $S_1x+S_2x=0$; similarly $S_1y+S_2y=0$, and hence $S_1d+S_2d=0$. Also if the number of equations is large S_1y and S_2y will be large, while as the d's are independent of the y's, S_1d will tend to contain as many negative as positive terms and to be small by comparison with S_1y: thus we might, when the number of values is large, take k as given by either of the equivalent equations

$$\left.\begin{array}{l} S_1\,x=k\,S_1\,y \\ S_2\,x=k\,S_2\,y \end{array}\right\} \quad \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \quad (2)$$

8. Such a process would however, when n is small, be open to an objection similar to that of para. 5 above. Thus if we consider, for simplicity, a case in which there are only four pairs of departures,

x	+1·03	—0·99	—0·98	+0·94
y	+0·51	+0·01	—0·49	—0·03

we note that the second and fourth pairs afford no reliable indications owing to the small value of y, while the first and third show that the variations of X are about double those of Y. Further if we interchange the first and second values of x, and second and fourth we do not affect S_1x and S_2x, or S_1y and S_2y: yet the numbers become

x	—0·99	+1·03	+0·94	—0·98
y	+0·51	+0·01	—0·49	—0·03

and the variations of X are now about minus twice those of Y, being in the opposite direction. The value of k as given by (2) is 1/13 in either case.

The incorrectness of these results obviously comes from failing to take into account the smallness of $+0{\cdot}01$ and $-0{\cdot}03$: or in other words attaching equal weight to each of the four equations

$$\left.\begin{aligned} +1{\cdot}03 &= +0{\cdot}51k + d_1 \\ -0{\cdot}99 &= +0{\cdot}01k + d_2 \\ -0{\cdot}98 &= -0{\cdot}49k + d_3 \\ +0{\cdot}9_{,} &= -0{\cdot}03k + d_4 \end{aligned}\right\}$$

It is obvious that the value of the indications given by those equations in which is small must be small to a proportionate amount, and hence that weight must be attached to the equations depending on the value of y; also the weight *may* be taken as proportional to that value, at any rate as a first approximation.

Thus the equations (1) will, on being weighted, become

$$\left.\begin{aligned} x_1 y_1 &= k y_1^2 + d_1 y_1 \\ x_2 y_2 &= k y_2^2 + d_2 y_2 \\ \ldots\ldots &= \ldots\ldots \\ x_n y_n &= k y_n^2 + d_n y_n \end{aligned}\right\}$$

and on adding we have

$$\mathrm{S}xy = k\mathrm{S}y^2 + \mathrm{S}dy \quad \ldots \quad (3)$$

Now since by definition the d's are independent of the y's there will when n is large tend to be as large negative contributions to $\mathrm{S}ly$ as positive contributions, and $\mathrm{S}dy$ will tend to vanish by comparison with $\mathrm{S}y^2$. Thus in the limit we shall have

$$\mathrm{S}xy = k\mathrm{S}y^2 \quad \ldots \quad (4)$$

9. As the departures of X, Y are negative as well as positive it is natural to define their mean values s, s_1 in terms of the squares of x, y by the equations

$$ns_1^2 = \mathrm{S}x^2,\ ns_2^2 = \mathrm{S}y_2^2$$

and it is convenient to introduce a new quantity r defined by the equation

$$nrs_1 s_2 = \mathrm{S}xy \quad \ldots \quad (5)$$

Thus (4) becomes

$$nrs_1 s_2 = kns_2^2$$

and $$k = rs_1/s_2 \quad \ldots \quad (6)$$

10. Let us now ascertain the proportionate extent to which the variations of X are determined by the variations of Y. Of the departures x_1, x_2, ... x_n the amounts which are determined by Y are ky_1, ky_2, ... ky_n, and of these the mean may be denoted by ms_1 where, by our definition of the mean,

$$\begin{aligned} n(ms_1)^2 &= (ky_1)^2 + (ky_2)^2 + \ldots\ldots + (ky_n)^2 \\ &= k^2\,\mathrm{S}\,y^2 = k^2\,n\,s_2^2 = n\,(rs_1)^2 \qquad \text{by (6)} \end{aligned}$$

Thus $m = r$.

Now the mean value of the variations of X is s_1, and the mean value of *those* portions which are dependent on Y is, as we have seen, ms_1 or rs_1. Thus r is the fraction of the variations of X which is determined by Y. Further from the definition (5) of r its value will not be affected if we replace all the x's by the corresponding y's, and hence it is also equal to the fraction of the variations of Y which are related to those of X. Thus it expresses the proportionate extent to which the variations of each are determined by, or related to, those of the other. It is usually called their correlation coefficient.

11. If the value of the d's be derived from equations (1) we shall have

$$Sd^2 = S(x-ky)^2 = n(s_1^2 - 2krs_1 s_2 + k^2 s_2^2)$$

and if we choose k so as to make Sd^2 a minimum we take $rs_1 = ks_2$, as in (6). Thus the value of k previously obtained makes the sum of the squares of the residuals or accidental differences d a minimum.

12. A similar method is applicable when we have three series of related quantities $X_1, X_2, \ldots\ldots X_n$; $Y_1, Y_2, \ldots\ldots Y_n$; and $Z_1, Z_2, \ldots\ldots Z_n$. As before we shall designate the departures of X, Y, Z from their average values by $x_1, x_2, \ldots\ldots x_n$; $y_1, y_2, \ldots\ldots y_r$; and $z_1, z_2, \ldots\ldots z_n$. Thus $Sx=0$, $Sy=0$, $Sz=0$: we also define the mean values s_1, s_2, s_3 by the equations

$ns_1^2 = Sx^2$, $ns_2^2 = Sy^2$, $ns_3^2 = Sz^2$ and quantities r_1, r_2, r_3 by the equations $nr_1 s_2 s_3 = Syz$, $nr_2 s_3 s_1 = Szx$, $nr_3 s_1 s_2 = Sxy \ldots\ldots$ (7).

We then assume that the variations of x are determined by those of y and z by equations of the form

$$\left.\begin{array}{l} x_1 = k_2 y_1 + k_3 z_1 + d_1 \\ x_2 = k_2 y_2 + k_3 z_2 + d_2 \\ \ldots \quad \ldots \quad \ldots \\ x_n = k_2 y_n + k_3 z_n + d_n \end{array}\right\} \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad .. \quad (8)$$

where $d_1, d_2, \ldots\ldots$ are independent of x, y and z; and we are unable to apply our previous analysis without modification because y, z are not independent of each other. We can however determine a constant c such that $z-cy$ shall be independent of y; the condition, as the previous work has shown, is that $y_1(z_1-cy_1) + y_2(z_2-cy_2) + \ldots\ldots y_n(z_n-cy_n)$ shall vanish, *i.e.*,

$$Syz - cSy^2 = 0 \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (9)$$

whence

$$nr_1 s_2 s_3 - ncs_2^2 = 0$$

or

$$cs_2 = r_1 s_3 \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (10)$$

Writing now the equations (8) in the form

$$\left.\begin{array}{l} x_1 = (k_2 + ck_3) y_1 + k_3(z_1 - cy_1) + d_1 \\ x_2 = (k_2 + ck_3) y_2 + k_3(z_2 - cy_2) + d_2 \\ \ldots = \quad \ldots \quad \ldots \\ x_n = (k_2 + ck_3) y_n + k_3(z_n - cy_n) + d_n \end{array}\right\} \quad \ldots \quad \ldots \quad \ldots \quad (11)$$

we note that all the terms $(z-cy)$ and d are independent of y, and our former analysis is applicable. We can thus deduce the value of (k_2+ck_3) by multiplying the equations (11) by $y_1, y_2, \ldots\ldots y_n$, omitting the d's and adding; we obtain

$$Sxy = (k_2 + ck_3) Sy^2$$

whence by (9)

$$Sxy = k_2 Sy^2 + k_3 Syz \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (12)$$

or

$$r_3 s_1 = k_2 s_2 + k_3 r_1 s_3 \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (13)$$

It will be seen that the equation (12) is what would be obtained if we had multiplied our original equations (8) by $y_1, y_2, \ldots\ldots y_n$, omitted the d's and added. In the same manner we may obtain the equation

$$Sxz = k_2 Syz + k_3 Sz^2$$

or

$$r_2 s_1 = k_2 r_1 s_2 + k_3 s_3 \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (14)$$

From (13) and (14) we find

$$\left.\begin{array}{l} k_2=s_1(r_2-r_1r_3)/\ s_2(1-r_1^2) \\ k_3=s_1(r_3-r_2r_1)/\ s_3(1-r_1^2) \end{array}\right\} \quad \ldots \quad \ldots \quad \ldots \quad \ldots (15)$$

The proportionate extent to which the variations of X are governed by those of Y and Z is the ratio to s_1 of the mean value of k_2y+k_3z, *i.e.*, the proportionate extent is m where $n\ m^2s_1^2=\mathrm{S}(k_2y+k_3z)^2$

$$=n(k_2^2s_2^2+2k_2k_3r_1s_2s_3+k_3^2s_3^2)$$

On substitution and reduction we obtain

$$m^2=(r_2^2+r_3^2-2r_1r_2r_3)/(1-r_1^2)$$

and the effective correlation coefficient of x with y and z is

$$\left\{(r_2^2+r_3^2-2r_1r_2r_3)/(1-r_1^2)\right\}^{\frac{1}{2}} \quad \ldots \quad \ldots \quad \ldots \quad \ldots (16)$$

13. The case of four variables may be treated similarly. If x, y, z, w be representative departures of four variables X, Y, Z, W from their average values, we shall define $s_1, s_2, s_3, s_4, r_{12}, r_{13}, r_{14}, r_{23}, r_{24}, r_{34}$ by the equations

$$\mathrm{S}x^2=ns_1,\ \mathrm{S}y^2=ns_2^2,\ \mathrm{S}z^2=ns_3^2,\ \mathrm{S}w^2=ns_4^2,\ \mathrm{S}xy=n\ r_{12}s_1s_2$$
$$\mathrm{S}xz=nr_{13}s_1s_3,\ \mathrm{S}xw=n\ r_{14}s_1s_4,\ \mathrm{S}yz=n\ r_{23}s_2s_3,\ \mathrm{S}yw=nr_{24}s_2s_4$$
$$\mathrm{S}zw=nr_{34}s_3s_4$$

Then as before we assume equations of the form

$$\left.\begin{array}{l} x_1=a_{12}y_1+a_{13}z_1+a_{14}w_1+d_1 \\ x_2=a_{12}y_2+a_{13}z_2+a_{14}w_2+d_2 \\ \ldots= \quad \ldots \quad \ldots \\ x_n=a_{12}y_n+a_{13}z_n+a_{14}w_n+d_n \end{array}\right\}$$

On multiplying by $y_1, y_2, \ldots\ldots y_n$, omitting the d's and adding we obtain

$$r_{12}s_1=a_{12}s_2+a_{13}r_{23}s_3+a_{14}r_{24}s_4 \quad \ldots \quad \ldots \quad \ldots \quad \ldots (17)$$

Similarly multiplying by $z_1, z_2, \ldots\ldots z_n$ gives

$$r_{13}s_1=a_{12}r_{23}s_2+a_{13}s_3+a_{14}r_{34}s_4 \quad \ldots \quad \ldots \quad \ldots \quad \ldots (18)$$

and by $w_1, w_2, \ldots\ldots\ldots w_n$

$$r_{14}s_1=a_{12}r_{24}s_2+a_{13}r_{34}s_3+a_{14}s_4 \quad \ldots \quad \ldots \quad \ldots \quad \ldots (19)$$

The equations (17), (18) and (19) are sufficient to give a_{12}, a_{13} and a_{14}.

Hence we find after some algebraic reduction

$$a_{12}=s_1\left\{r_{12}(1-r_{34}^2)+r_{13}(r_{24}r_{34}-r_{23})+r_{14}(r_{23}r_{34}-r_{24})\right\}$$
$$\div s_2\left\{1-r_{23}^2-r_{24}^2-r_{34}^2+2r_{23}r_{24}r_{34}\right\}$$

with corresponding values for a_{13}, a_{14}, &c.

Defining as before the correlation coefficient m of x with y, z, w, as the proportionate extent to which the variations x are determined by y, z, and w we have, as before,

$$m^2s_1^2=a_{12}^2s_2^2+a_{13}^2s_3^2+a_{14}^2s_4^2+2a_{12}a_{13}r_{23}s_2s_3+2a_{12}a_{14}r_{24}s_2s_4+2a_{13}a_{14}r_{34}s_3s_4 \quad \ldots \quad (20)$$
$$=a_{12}s_2r_{12}s_1+a_{13}s_3r_{13}s_1+a_{14}s_4r_{14}s_1$$

by (17), (18) and (19),

$$=s_1(a_{12}r_{12}s_2+a_{13}r_{13}s_3+a_{14}r_{14}s_4)$$

Thus

$$m^2\begin{vmatrix} 1, & r_{23}, & r_{24} \\ r_{23}, & 1, & r_{34} \\ r_{24}, & r_{34}, & 1 \end{vmatrix} = \frac{-1}{s_1}\begin{vmatrix} a_{12}r_{12}s_2+a_{13}r_{13}s_3+a_{14}r_{14}s_4, & 0, & 0, & 0 \\ r_{12}, & 1, & r_{23}, & r_{24} \\ r_{13}, & r_{23}, & 1, & r_{34} \\ r_{14}, & r_{24}, & r_{34}, & 1 \end{vmatrix}$$

On multiplying the 2nd, 3rd and 4th rows by $-a_{12}s_2, -a_{13}s_3, -a_{14}s_4$, and adding to the first, this becomes, on using (17), (18) and (19):—

$$\frac{-1}{s_1}\begin{vmatrix} 0, & -r_{12}s_1, & -r_{13}s_1, & -r_{14}s_1 \\ r_{12}, & 1, & r_{23}, & r_{24} \\ r_{13}, & r_{23}, & 1, & r_{34} \\ r_{14}, & r_{24}, & r_{34}, & 1 \end{vmatrix}$$

or

$$-\begin{vmatrix} 0, & r_{12}, & r_{13}, & r_{14} \\ r_{12}, & 1, & r_{23}, & r_{24} \\ r_{13}, & r_{23}, & 1, & r_{34} \\ r_{14}, & r_{24}, & r_{34}, & 1 \end{vmatrix}$$

Hence m is given by

$$(1-m^2)\begin{vmatrix} 1, & r_{23}, & r_{24} \\ r_{23}, & 1, & r_{34} \\ r_{24}, & r_{34}, & 1 \end{vmatrix} = \begin{vmatrix} 1, & r_{12}, & r_{13}, & r_{14} \\ r_{12}, & 1, & r_{23}, & r_{24} \\ r_{13}, & r_{23}, & 1, & r_{34} \\ r_{14}, & r_{24}, & r_{34}, & 1 \end{vmatrix} \quad \ldots \quad \ldots \quad (21)$$

14. Another theorem follows from noticing that the equations (17), (18) and (19) are equivalent to

$$Sdy=0, \qquad Sdz=0, \qquad Sdw=0$$

and hence

$$Sd\,(a_{12}y+a_{13}z+a_{14}w)=0 \quad \ldots \quad \ldots \quad \ldots \quad (22)$$

i.e., $\quad Sd\,(x-d)=0$

Thus $\quad S\,d^2=S\,dx \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (23)$

But $\quad d_1^2=x_1^2-2x_1\,(a_{12}y_1+a_{13}z_1+a_{14}w_1)+(a_{12}y_1+a_{13}z_1+a_{14}w_1)^2$

and hence summing the similar equations

$$Sd^2=Sx^2-2Sx\,(x-d)+m^2S\,x^2$$

where the last term follows from the definition of m.

Thus by (23)

$$Sd^2=Sx^2-2Sx^2+2S\,d^2+m^2Sx^2$$

i.e. $\quad S\,d^2=(1-m^2)\,Sx^2$

Thus the mean value of d is $s_1\,(1-m^2)^{\frac{1}{2}}$, $\quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (24)$

and the more closely the correlation coefficient m approaches unity the smaller is the mean value of d.

15. It is natural after trying to explain the variations of X in terms of those of Y, Z and W in any climatic problem to evaluate the quantities

$$\left.\begin{array}{l} a_{12}y_1+a_{13}z_1+a_{14}w_1 \\ a_{12}y_2+a_{13}z_2+a_{14}w_2 \\ \ldots \quad \ldots \quad \ldots \\ a_{12}y_n+a_{13}z_n+a_{14}w_n \end{array}\right\} \ldots \quad \ldots \quad \ldots \quad \ldots \qquad (25)$$

which are the values of x that would be inferred from known values of the y's, z's and w's. After evaluating these it is natural to work out their correlation coefficient with the actual values of x. The sum of the squares of the quantities (25) is, by the definition of m, equal to $nm^2s_1^2$, and hence the correlation coefficient of the series (25) with the x's is

$$S\,x\,(a_{12}y+a_{13}z+a_{14}w)\,/\,ns_1ms_1$$

or

$$S[(a_{12}y+a_{13}z+a_{14}w)+d](a_{12}y+a_{13}z+a_{14}w)\,/\,mns_1^2.$$

Now in the numerator

$$S\,(a_{12}y+a_{13}z+a_{14}w)^2=n\,m^2s_1^2$$

and, by (22)

$$Sd(a_{12}y+a_{13}z+a_{14}w)=0.$$

Hence our fraction becomes $nm^2s_1^2 \div mns_1^2$, or m, the correlation coefficient of x with the variables, as defined by the proportionate extent to which its variations are governed by those of y, z, and w.

16. Another consequence of (17), (18) and (19) is easily seen to be that the coefficients a_{12} &c., are so chosen as to make Sd^2 a minimum. The results of this and the three previous paragraphs are obviously applicable to any number of variables.

Result of assuming the exponential law of distribution in the variations.

17. The values of the correlation coefficients have been obtained by making the simplest hypotheses. The equations (17), (18) and (19) follow at once from assuming the exponential law of distribution, and Pearson has shown that when that law holds the probable error of the coefficient r of correlation between two variables is $\cdot 67449(1-r^2)/\sqrt{n}$ where n is the number of pairs of values.* This result will frequently be utilised in the succeeding papers on this subject as an approximation when the distribution law of variation does not differ far from the exponential law.

* Philosophical Transactions, London, Vol. 191, p. 242 (1898).

CALCUTTA: PRINTED BY SUPDT. GOVT. PRINTING, INDIA, 8, HASTINGS STREET.